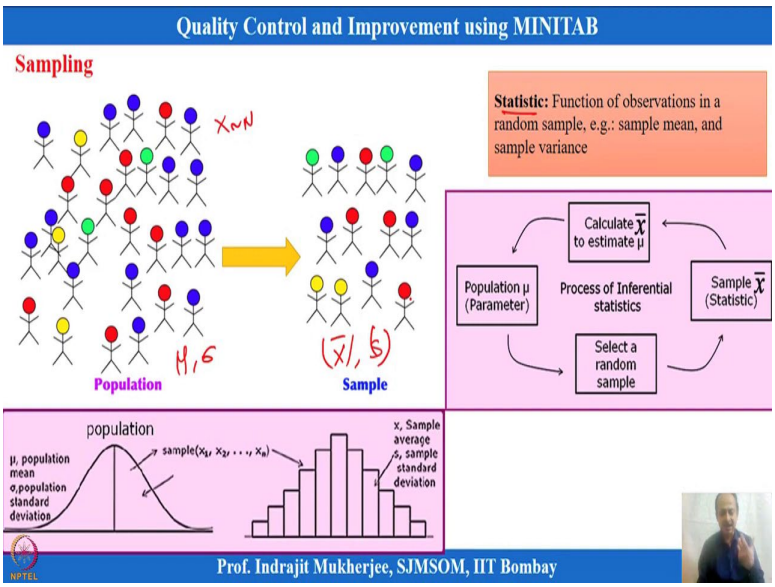
Quality Control and Improvement with MINITAB
Prof. Indrajit Mukherjee
Shailesh J. Mehta School of Management
Indian Institute of Technology, Bombay

Lecture - 16
Basic Statistics and Confidence Interval

Hello, welcome back to session 16 on Quality Control and Improvement with MINITAB. I am Professor Indrajit Mukherjee from Shailesh J. Mehta School of Management, IIT Bombay. So, in the previous session, what we are doing is that we are giving some basic ideas on statistics which will be used for quality control and improvement.

So, I will only touch the relevant part, which is required for our course and we will not go in to much details about, because that can be you already have a sense of that in statistical course or any other course on basic statistics like that ok.

(Refer Slide Time: 00:49)



But, I will give you a hint what we are doing over here for in quality control is that, we take samples let us say machine outputs are coming and in that case, what we do is that infinite populations we can assume for that. And, in that case we take some sample and that will be giving us some snapshot about the process and based on the what we do is that we try to say that this is the process capability and in short term process capability this is the long term process capability.

So, because I do not have population information, in that case some samples we or reasonable samples we collect and based on that we try to infer about the population. So, over here whenever I am taking a sample, I estimate some values over here that is statistic what we call, maybe average value what we are interested into and may be also we are interested into standard deviation of the samples that we have taken.

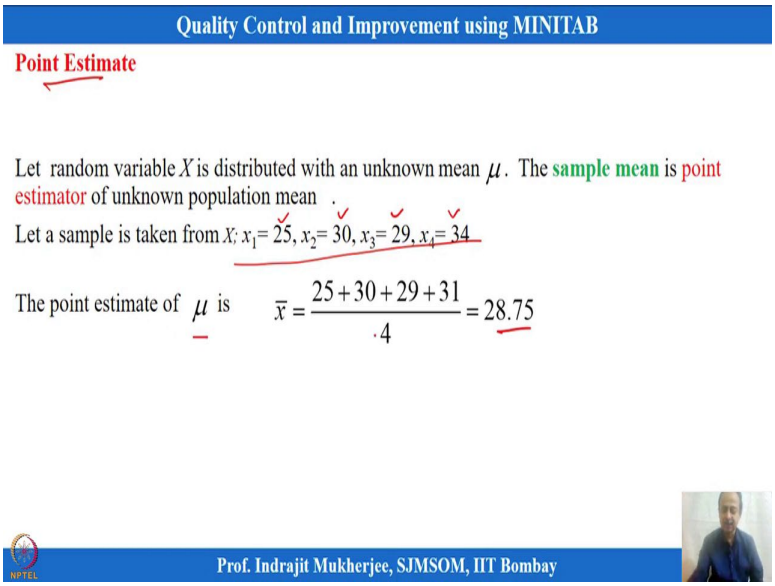
And, based on that we want to predict what will be the μ of the population and what will be the σof the population; that means, standard deviation of the population like that. So, this X-bar s is known as statistic this X-bar s is known as statistic and this statistic can follow certain distributions like that.

This can also follow certain distribution like, individual this if we assume this is the target population and these are the random variables X, that can follow normal distribution over here and over here statistic that values that we are taken, also can follow average value that we select, because if I take a different sample it will be a different X-bar that we get out of the population like that.

So, next time I go to the process, I take some 10 samples like that, I will get a different estimation of X-bar like that. So, X-bar every time we calculate what we will get is that every time it will be different. So, X-bar can also be considered as a random variable, because initial X values that we have selected and that are selected based on randomness; that means, we have assumed that there is no bias in the sample selection like that ok.

So, every time I am selecting the samples so, that is giving me a value of X-bar and S and these values can change, if I change the samples like that if I continuously take more samples it will be different. So, this is known as statistic that we are measuring over here and this statistic is also a random variable here.

(Refer Slide Time: 03:04)



And so, why I am doing this, because I want to predict what is the behavior of the population because I do not have access to population. So, I am going to predict that one what will be the behavior of population. So, one of the important statistical concept is over point estimation of the statistic that we are doing.

So, if I have four samples like that, let us consider an apple tree in that case you have taken four samples and based on that you calculate the average. And you say, this average I am expecting the population tree. It is unlikely that it will be exactly equals to that, but this is the estimation I can make and I have taken the sample unbiasedly.

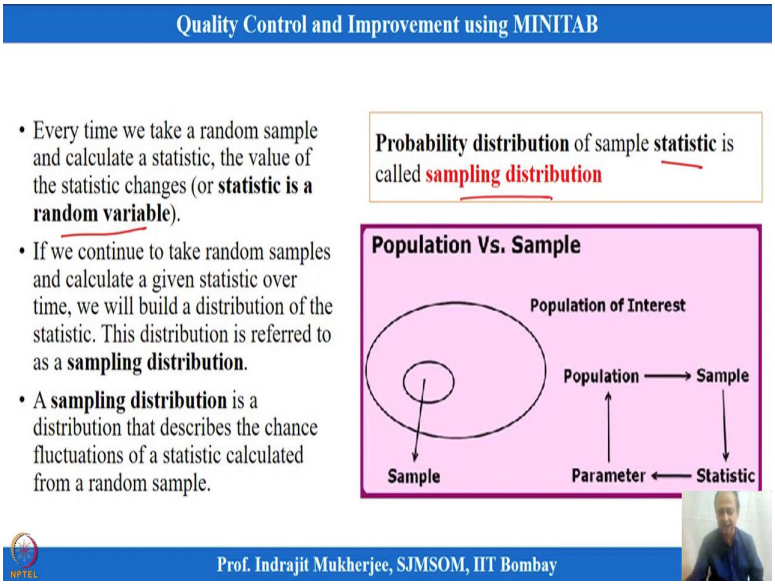
So, in that case I am predicting and we say that this is a point estimate of the population μ over here. So, in this case although it cannot be exactly equals to μ because, it is rear rarest possibility that the average that you have calculated will be exactly equals to μ, ok.

But, I make an estimation over here. So, this estimation what we are making about the, we are trying to predict the population mean over here. So, one of the estimation is point estimation that we are making and we are just saying that this $\bar{x}$ will be representative of the population μ like that ok. So, this is known as point estimation.

So, when I am making point estimation about the population parameter. So, here $\bar{x}$ is calculated. So, it is calculated as one of the observation is 25, 30, 29 and 34. So, average

of this is the point estimation that we are getting for the population μ over here so, estimation parameter μ over here ok.

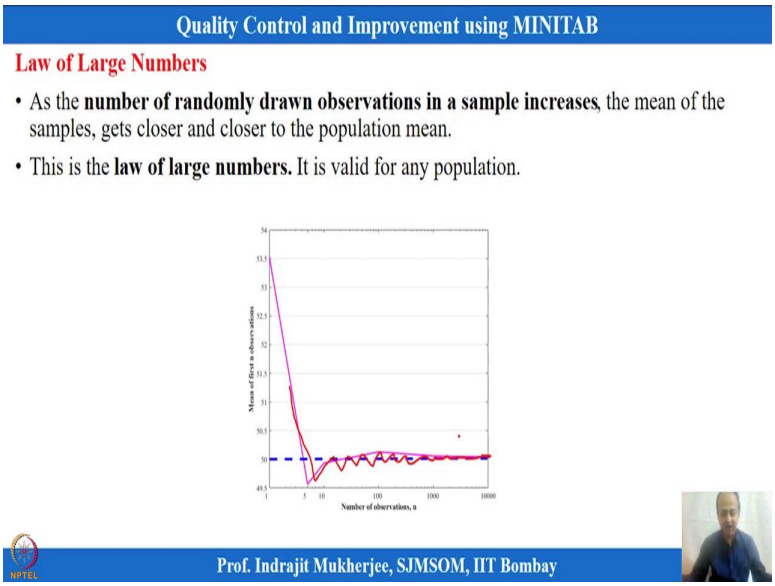
(Refer Slide Time: 04:36)



Now, this population parameter will keep on change like that it will differ, because every time I take a samples it will be different. So, this will also be a random variable and this population these sample parameters can. If it is a random variable it will follow certain distribution, it can follow certain distribution or maybe popular distribution like normal distributions like that ok.

So, any other distribution we can think of. So, this statistic can follow certain sampling distribution. So, probability distribution of a statistic is known as sampling distribution, because every time I am taking a samples and based on that $\bar{x}$ let us say every time I am calculating and $\bar{x}$ will follow certain distributions like that ok.

(Refer Slide Time: 05:17)

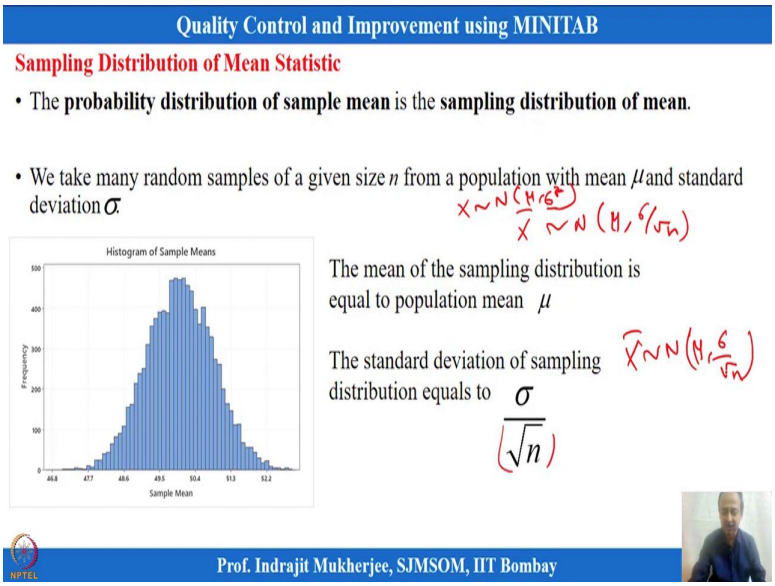


So, that is the idea of sampling distribution. So, I am doing sampling distribution and also one important theory that comes into that is also important over here, which is known as law of large number; that means, if you have large number of observations and the mean of the samples gets closer and closer to the population mean like that.

So, over all mean of the population, if you take every time some means some and you keep on drawing the samples like that it will happen that this fluctuation over here. So, initial samples that we have taken and the fluctuation will be converging to the population mean like that.

So, more and more you take independent observations average values and average will tend towards the population average like that. So, that is known as law of large number ok.

(Refer Slide Time: 05:56)



And, when we do sampling distribution of the means over here that is when we calculate the sample when we try to determine the probability distribution of the sample means statistician has given us a formulation over here. They it says that $\bar{x}$ we will follow normal over here with mean will be same as original X follows.

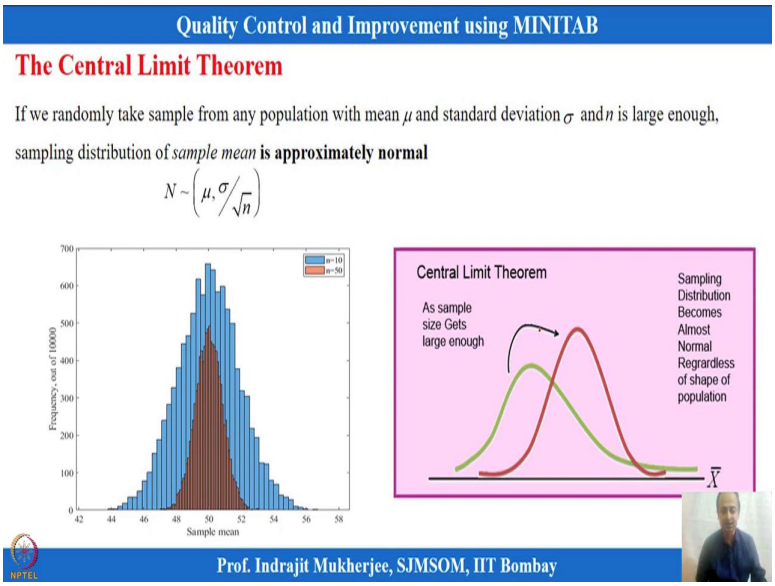
Let us say x follows normal over here and with μ and σ^2 over here. So, $\bar{x}$ will follow normal with μ and σ^2 / n like that.

So, over here what happens is that distribution is somewhat more, we can say more closer as compared to the original x populations like that this or the variability of the variability of the statistic will decrease, because here we are taking a factor of $\sqrt{n}$ in the denominator like that. So, earlier this is sigma. So, σwill reduced and σwill be for the sampling distribution it will be σby $\sqrt{n}$ over here.

So, this is one of the things that we have to consider, when we are analyzing afterwards using statistics like that $\sigma / \sqrt{n}$ is the standard deviations, estimated standard deviation of the sample statistics ok, sample statistic which is over here average value; that means, $\bar{x}$.

So, $\bar{x}$ follows normal with mean remain same and only σchanges to the σby root n like that, because I am taking average one smoothing it out in that case fluctuations will be less, variability will be less. So, $\sigma / \sqrt{n}$ is the formulation that statistician has given us ok.

(Refer Slide Time: 07:40)

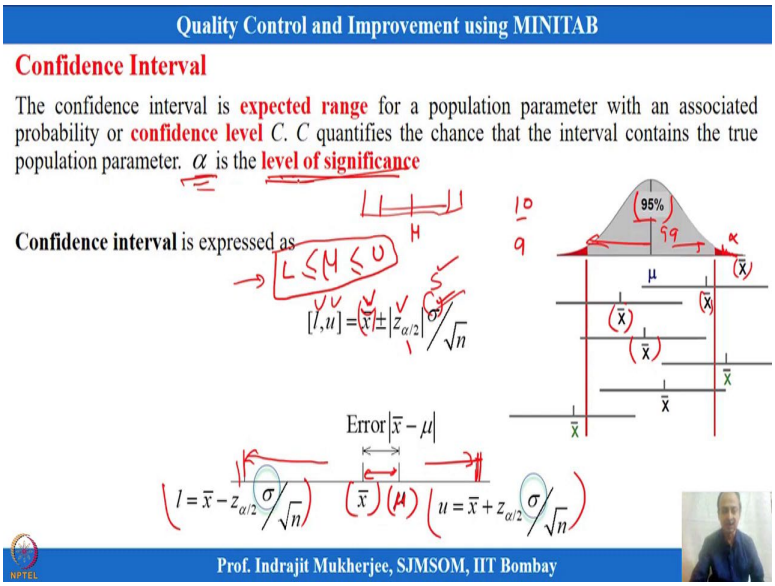


And central limit theorem also is another important concept that is used; that means, a underlying X variables can be followed any types of distribution, but if you take the average like sample statistics over here, average over here is considered. So, in that case what will happen is that that will converge to normal distributions like that.

It can be of any distribution, it can be proved like that as you increase the sample numbers and you take the average and then try to plot the average, what will happen is that the distribution of the average will follow normal like that. So, that is why what you see in quality is that we take more number of observations to calculate the average and we assume that average will tend towards normal like that, theory says like that way ok.

So, like in control chart we take subgroups size what you have seen like that. So, five subgroups and I take the average and we plot the average like that and we assume the average follows normal distribution. So, basic underlying idea over here is the central limit theorem that we can consider over here ok.

(Refer Slide Time: 08:35)



And, then another important concept that is also required, when we move forward is known as confidence interval over here. So, confidence interval idea says that exactly I cannot heat whenever I am calculating some samples from the apple tree and I am trying to predict the population average values of the weight of the apples like that.

So, in that case what will happen is that I can only do destructive stressing or takes 4 or 5 apples and I cannot do it several times like that that is uneconomical for me. So, in that case what I will do is that I will take one samples and based on that I would like to predict what where will be the population average values like that.

So, population average, so, for that one concept was developed over here by the statistician which is known as confidence interval, over here with one estimation of X-bar what we can say is that with certain confidence that the lower bound and upper bound where μ will lie I can determine over here.

So, this L and and U, let us say lower and upper limits like that. So, some bounds can be given to this μ based on one estimation which is X-bar over here. So, I have only one estimation and I have some information on standard deviation over here. See if I have mean information is standard deviation information which can be also sample standard deviation over here.

So, if you have these two parameter information or some estimates over here, then I can say where the μ should lie; that means, where the μ should lie. This confidence interval and this gives you better assessment over here, rather than saying $\bar{x}$ will be exactly equals to μ. I am saying that if $\bar{x}$ is this much, μ should lie within L and U, these two this values like that.

So, I am giving a confidence interval over here and confidence interval over here depends on certain value which is known as z statistics over here, and assuming that the σvalues of the population is known. Most of the time it may not be known, so, in that case this z will be replaced by t-statistic like that.

So, there are different distributions, continuous distribution very some of the important distribution used in quality for assessing and design of experiments are F-distribution, t-distribution and Z-distribution that is normal distributions like that ok.

So, normal distribution, t-distribution, F-distribution these are the common distributions or chi-square distribution. So, these are the distribution which will be used for quality analysis like that quality control and improvement analysis like that ok.

So, one of the important concepts over here is shown over here. So, μ will lie. So, if you take an average, one average over can fall over here, one average can fall over here, one average can fall over here, but whenever you are building the confidence interval you can be sure. So, let us say they have developed a 95 percent confidence interval and I have one average. It is expected that and you d multiple times, I take multiple average over here. So, if I have taken 10 times I have taken the sample. So, I can expect that over here 9.5 times so, over here. So, if it is 90 percent over here, we can assume that 9 out of 10 or more than that will be within the confidence interval that I have given over here, so, in this case.

So, that is given as confidence interval like that. So, one sample with some confidence level which is given over here as α over here that is known as α is known as level of significance, α is known as level of significance and if you can define the α what level of significance is one that will define the width of this confidence interval over here that will define the width of the confidence interval over here.

So, if you want to be more confident over here, you increase the α values over here and in that case the calculation will show you give you a wider range of this. So, that means, you can expect that if I increase the increase the level of confidence, if I have to increase the level of confidence in that case α has to be decreased over here.

So, in this case error or committing error which is known as α over here. So, this α values needs to be decreased over here. You want to be more confident over here, so, you need to expand this band over here rather than 95, I want to expand these to 99 bandwidth of this.

And, in that case what will happen is that this area will go down over here and α will come down; α is the rejection basically, how much I can be wrong like that, how much time you can think of that probability of going wrong like that ok. My estimation of confidence interval can go wrong in what is the percentage chance of that like that ok.

So, if you define α and you calculate one average over here and you can calculate the standard deviation either by s or σover here, I can define a confidence zone or lower limit and upper limit over here I can define over here; that means, with a given estimation of X-bar which may not be there can be error between $\bar{x}$ and μ over here.

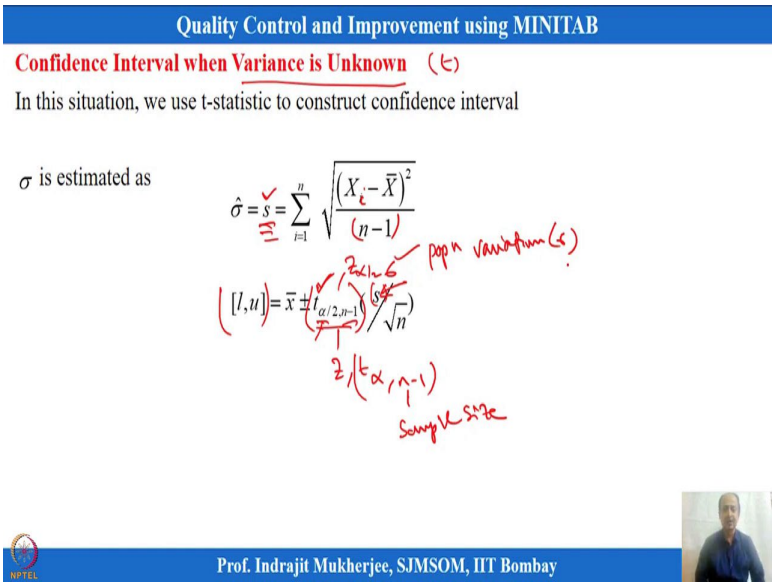
But, what we expect is that maximum whether I can commit when I am over here in this zone or in this zone. So, I am saying that μ is expected to lie within this zone and this zone over here. This is a confidence interval confidence, interval of μ over here or population μ over here ok.

So, I am giving a with some confidence over here. So, chances is that 95 percent of the time I will be right, but 5 percent of the time I can be wrong over here that is why probability is used over here. So, this is a concept of confidence interval we can think of.

So, this idea of confidence interval, so, if I take an average, so, what do you have to do is that you take samples from the process and based on that I want to infer about the population mean or standard deviation whatever you can think of. So, in that case statistician is given us some formulas that if this is the mean or this is the standard deviation, I can calculate what is the confidence interval of population parameter, where the population parameters is expected with certain confidence level.

That means, I can be wrong, but that will be defined by α levels what we are seeing as level of significance over here. So, you define the level of significance and give me X-bar values and estimation of variance over here, then I can tell you what is the confidence interval within which μ is expected or the or the population parameter is expected like that. So, that is known as confidence interval, that idea is given as confidence interval over here ok.

(Refer Slide Time: 14:45)



With that estimation, so, even σ can be estimated over here, unbiased estimation is s that is sample standard deviation over here again with this estimation, if you have S again lower and upper limits or bounds of the μ can be given over here, which is given in formulation and only t statistics is used over here, where α is the level of significance that I already I have mentioned. What is the chance that I can go wrong like that.

And, there will be some degree of freedom which is mentioned over here as n. So, these two if you can define and I can define what is the value like z values over here, we can also define what is the t value, for a given level of α and given level of degree of freedom I can always assume and this depends on the sample observation that you have taken. So, sample size (n -1) so, this will give me the estimation of n over here.

So, I can define the confidence interval over here. Only thing is that instead of σ I have used S over here. So, I am using σ in that case this over here, it will be replaced by Zα/2

over here and if S we have no estimation of σ or population variance or variation population variation or standard deviation over here which is expressed as σ.

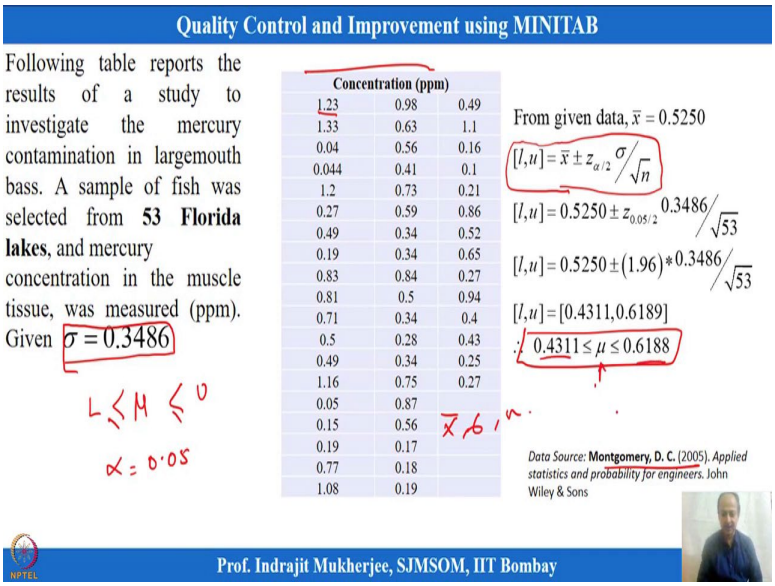
So, in that case what is possible is that I can just place the unbiased estimation which is s. s is a sample standard deviation over here which is calculated by individual observation by X-bar divided by n minus one as the degree of freedom. So, in this case what will happen is that I can replace that one only thing I will use a t-distribution instead of Z-distribution like that ok.

So, when variance is unknown. So, in that scenario t statistics will be used in that scenario t will be used to find out the confidence interval to find out the confidence interval. So, we have to remember two important things over here: one is known as confidence interval one is known as level of significance; that means, what is the probability that I can be wrong.

So, this will be defined by level of significance like that. If it is 5 percent, 5 percent of the time this confidence interval that I have given can be wrong like that. So, that is the, that is an interpretation I can make out of this ok confidence interval. And, if variance is known in that case we use z to define the bounds and if σif variance is unknown, population variance is unknown in that case it will be replaced by t statistic over here.

So, t it and we can get this t value from tables which is provided in any statistical book at the end of the books you will find, but MINITAB will do it automatically for you. So, if we define the samples and MINITAB will do it automatically for you.

(Refer Slide Time: 17:19)



So, this is one of the example that we will use to demonstrate a MINITAB does and builds the confidence interval like that. So, this is a table which shows investigation of mercury contamination is largemouth bass over here. So, a sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured like that in PPM and this is the measures that you see concentration over here, these are the values that we are getting.

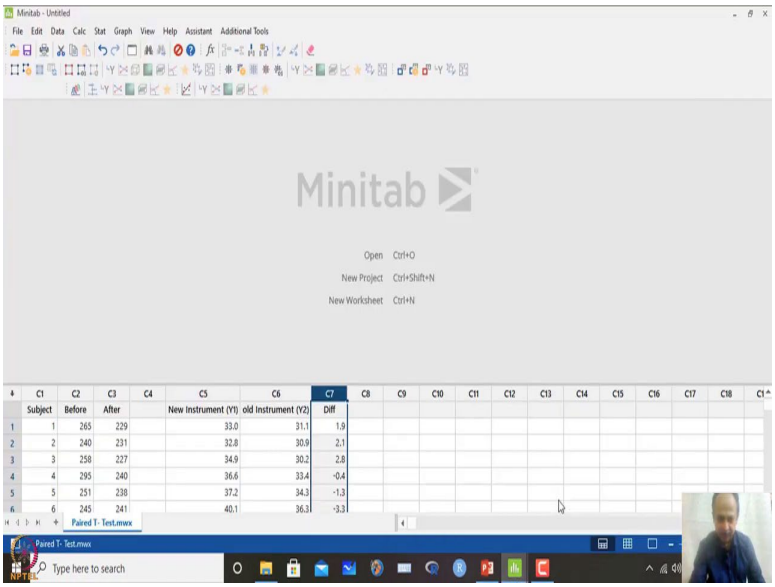
And, this data set is taken from Montgomery's book and the population standard deviation value is given over here. So, I want the mean upper bound and lower bound over here, so, lower bound. So, I want to build the, I want to determine the confidence interval of this. So, upper bound and lower bound over here.

So, in these case let us assume that α equals to 0.05 like that and that is a probability I can go wrong. So, with that what we can do is that, we can define the formulation over here which is given in the last slide also and this is the formulation for lower bound and upper bound calculation like that.

So, this is the data set that we are having, we can always calculate X-bar average over here and σ is known over here. So, in this case immediately and n is also known number of observations over here is given. So, I can calculate what is a lower limit even of I use the formulas.

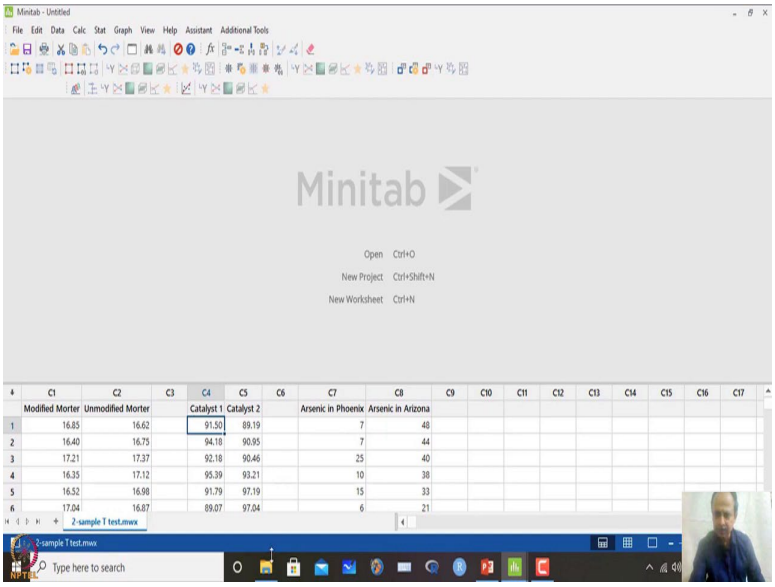
And, the final calculation what you see is that the confidence interval of μ is coming out to be 0.4311 to 0.6188 like that. Let us do it in MINITAB and try to do that whether MINITAB is also giving the same results or not ok.

(Refer Slide Time: 18:55)



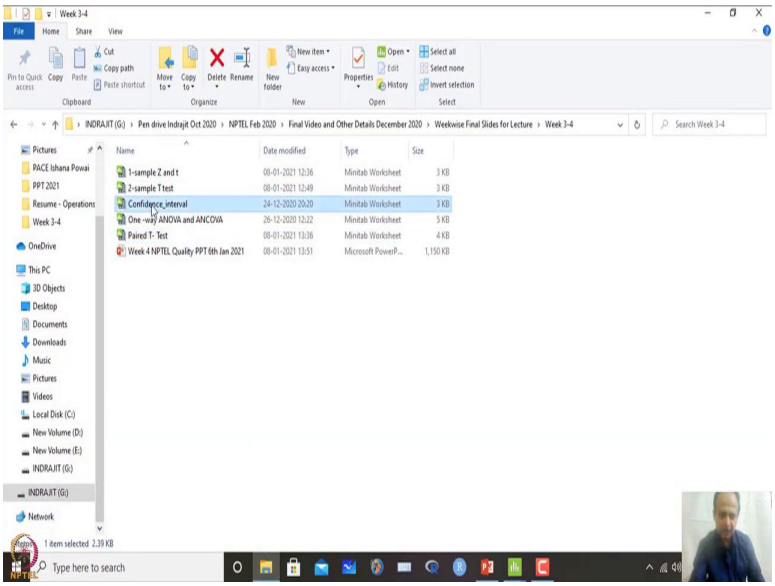
So, what I will do is that I will go to Minitab file where it is given.

(Refer Slide Time: 18:59)



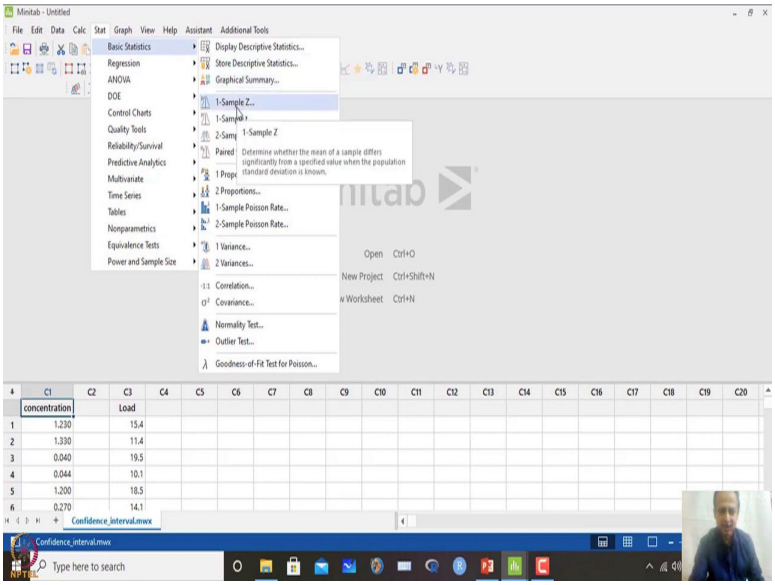
So, in this case what we will do is that. So, other than this file, we may be having another file.

(Refer Slide Time: 19:03)



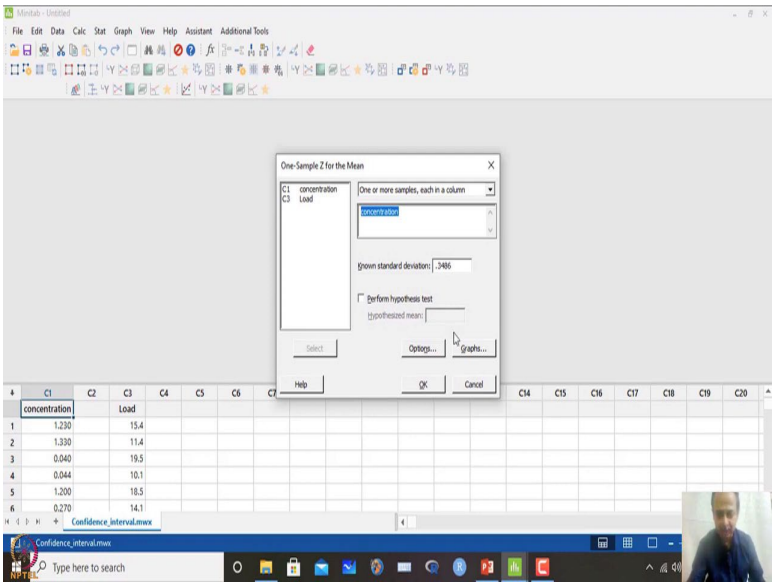
Where we have Confidence interval and the dataset will be there. So, let us open the dataset.

(Refer Slide Time: 19:11)



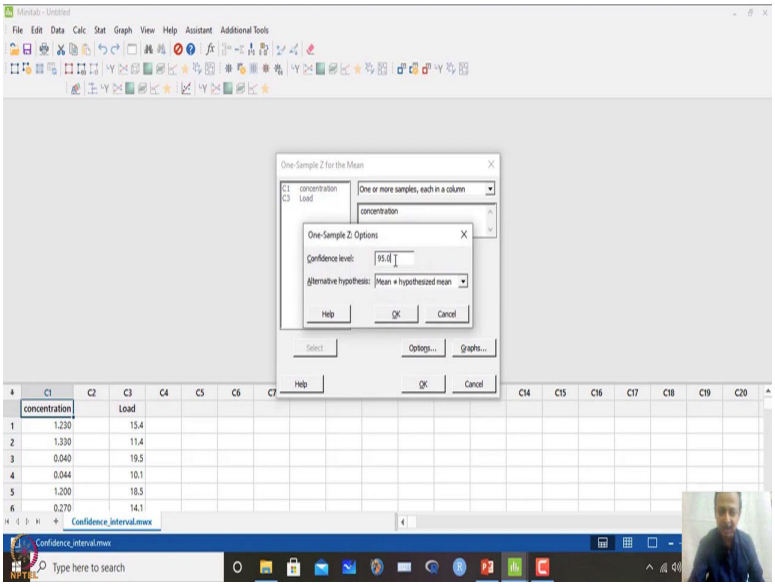
And this is the concentration what example we are talking about concentration in ppm. So, in this case what is given is that this data set is given over here. So, this is the concentration dataset that we are having and we want to find the confidence interval. So, when σis given. So, I will go to stat, Basic Stat maybe and I will go to 1-Sample Z over here.

(Refer Slide Time: 19:39)



So, in this case what I will do is that I will identify each in column. So, it is not summarize. So, concentration over here and known standard deviation I have to give over here. So, known standard deviation which is given is 0.3486 and that I will write 0.3486. And if I and then in that case everything I do not want to.

(Refer Slide Time: 20:00)

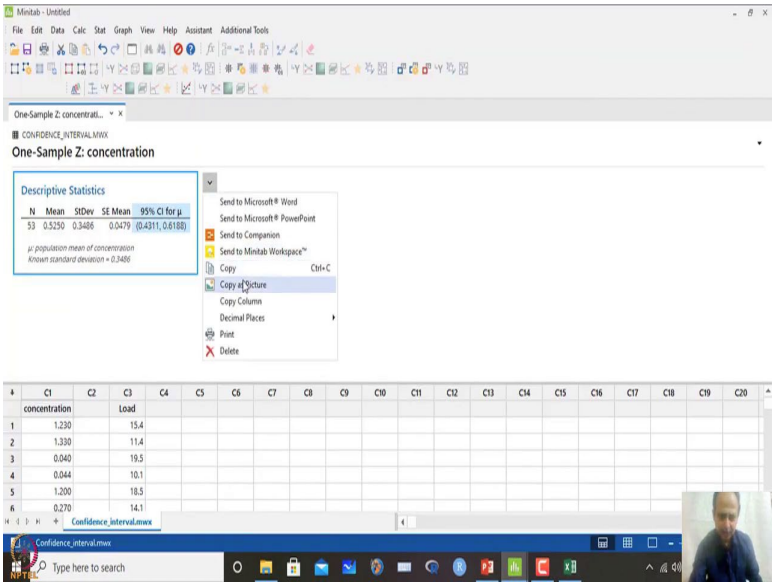


And, Confidence Interval that over here, one option is there 95 percent band you want or whatever you want that you have to mention. Now, this hypothesis testing we have not

discussed. So, ignore this one, do not click this one. So, I am not clicking anything over here.

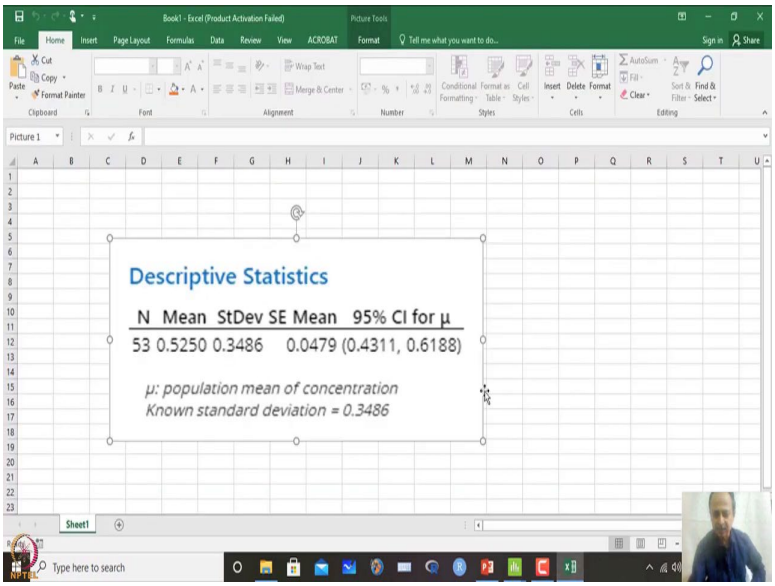
So, I am just trying to figure out, if σis known and the dataset is given where the population μ will lie like that. So, that is a confidence zone I want to calculate and for that in options what I have taken is that 95 percent confidence level that is confidence level is taken. So, α will be (100-95); that means 5 percent over here. So, that means, I can be wrong 5 percent over time. So, if you click Ok over here, and you click Ok.

(Refer Slide Time: 20:40)



What will happen is that you will get this information like this. So, in this case what happens is that, we can we can just enlarge this one. We can copy this one and we can also paste it in Excel, let us say and we can just see where the values may be it is from here it is difficult to see. So, I am just copy pasting this one in Excel sheet over here.

(Refer Slide Time: 21:08)



So, if I do that so, in this case what we have done is that so, once again just go back to the previous one. So, I will copy as picture if I do that and paste it over here. Now, it is possible. So, this is not required and I can just enlarge this one like that. So, you can now, it should be visible somewhat distorted, but this is visible over here.

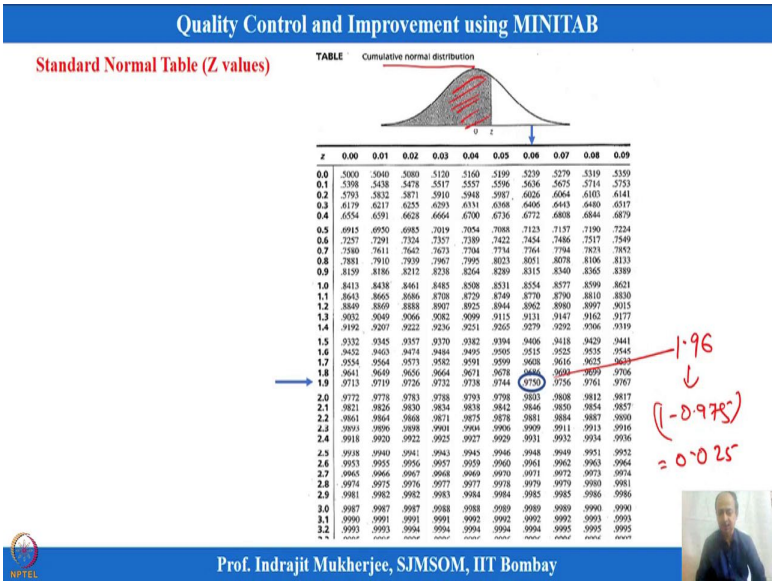
And, number of observation is 53 over here, mean of the observation is 0.52 that is the sample observation and sample standard deviation is given and standard error of mean over here is calculated by standard deviation divided by square root of n like that. So, you can calculate that one you will get standard error of mean that is the variability of the mean basically.

So, then 95 percent confidence interval of μ that is calculated is 0.4311 and my calculation hand calculation also says 4311 is a calculation and 6188 is the calculation when I am doing this. So, 6188 is also same. So, Minitab has calculated the same thing; if I have done it by hand also I am getting the same values over here.

So, known standard deviation is equals to 0.3486 that I have that I have to be input over here. So, in this case it becomes easier for me to calculate the confidence interval ok. So, ok, so, this is how we are calculating over here and confidence interval is very easy. So, I can calculate the confidence interval for a given sample observations.

So, similarly, what we can see now these Z values of α/2 what you see over here α/2. So, this is 0.025 over here. So, in this case what happens is that I need a Z table for determining for these Z statistics that is written over here $Z_{0.25}$. So, this value that you see over here is I also need over here. So, n is known, σis known, X-bar is known, but Z α /2 is not known to me.

(Refer Slide Time: 22:46)

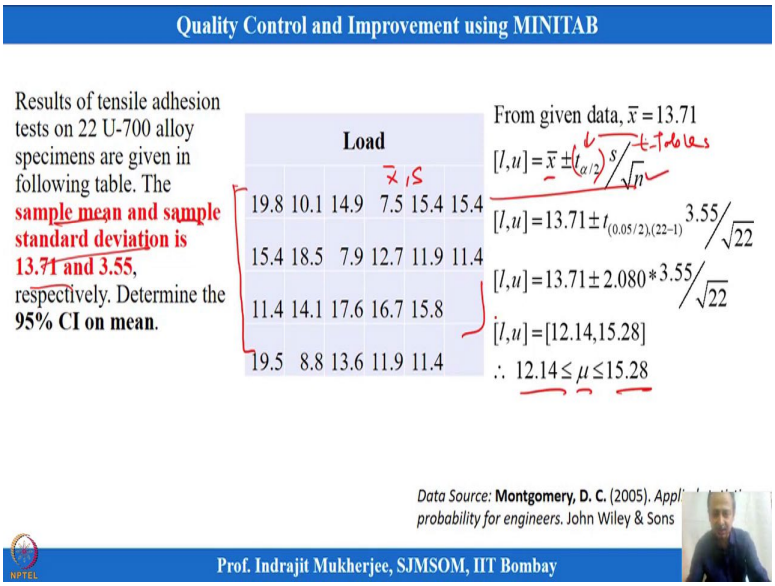


So, how do I get that Zα values over here? So, I have a standard normal table like that and standard normal table will tell me that where it is so, 9725. So, if you if your if you have done some basic course on statistics you must be knowing that how to see standard normal distributions like that.

So, in this case 1.96 this value is basically giving Z value of 1.96 that will give you a area under the curve; that means, area under the curve over here. So, this is the area shaded area over here, this is coming out to be 0.975 and 1 minus of 0.975 will give you 0.025 like that ok. So, this is the value that we are looking for and.

So, Z values can be seen from the Z tables like that and from there we can calculate this one. So, if Z is known then everything is known over here. So, then lower bound and upper bound calculation is easy. So, based on that I can calculate what is a lower bound and upper bound like that ok.

(Refer Slide Time: 23:59)



So, in case you do not have this standard deviation estimation like that. So, that means, only sample mean and sample standard deviation calculation is given. So, instead of Z what we need is that t value over here. So, this is given by statistician that to build the confidence interval in that case Z distribution cannot be used. So, t distribution has to be used.

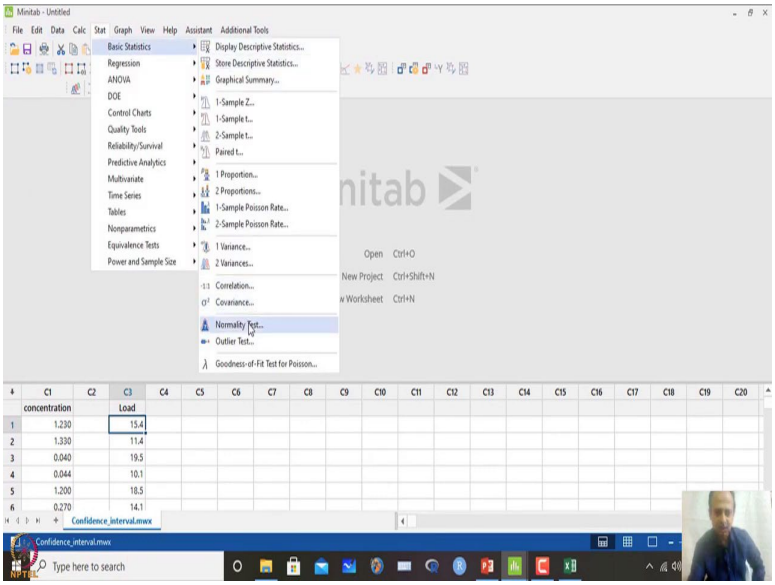
So, t is a specific distribution which is coming from Z distribution. So, this t distribution concept has to be used over here to determine the confidence interval over here, but underlying assumption is that this data follows normality, because this data follows normal over here. So, this assumption should be verified, and then only we can give the confidence interval over here.

So, in this case confidence interval using t statistics is coming out to be this and this can be calculated this can be seen from tables, t tables are there. So, t tables if you see so, in that case we will get the values of this. So, X-bar is known, standard deviation of this is known, X-bar is known. So, in that case immediately I can determine, n is also known over here.

So, how MINITAB does it gives for you? Let us go to MINITAB over here and let us try to see this and you do not have to do anything over here and what you have to do is that we will use the same one and this is the second dataset sorry, this is the second dataset we are using which is on loads. So, tensile adhesion tests that is given; so, this load is the

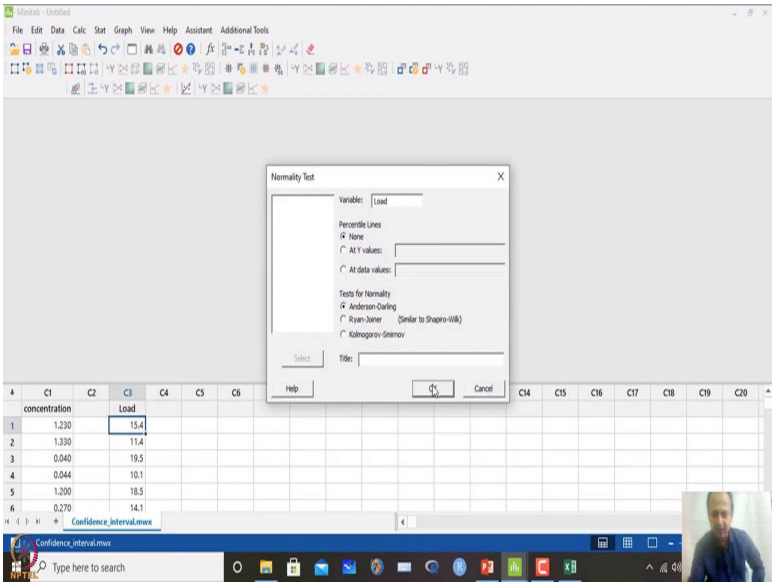
dataset that we are trying to see and we can go to the dataset and I told you how to see normal distribution assumption.

(Refer Slide Time: 25:27)



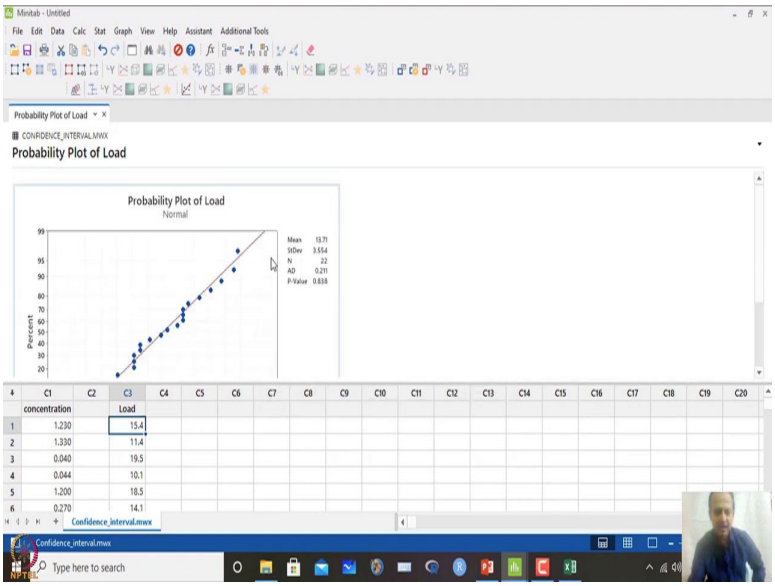
So, immediately we can verify. So, go to Basic Statistics, go to Normality Test and go to Load variables over here.

(Refer Slide Time: 25:31)



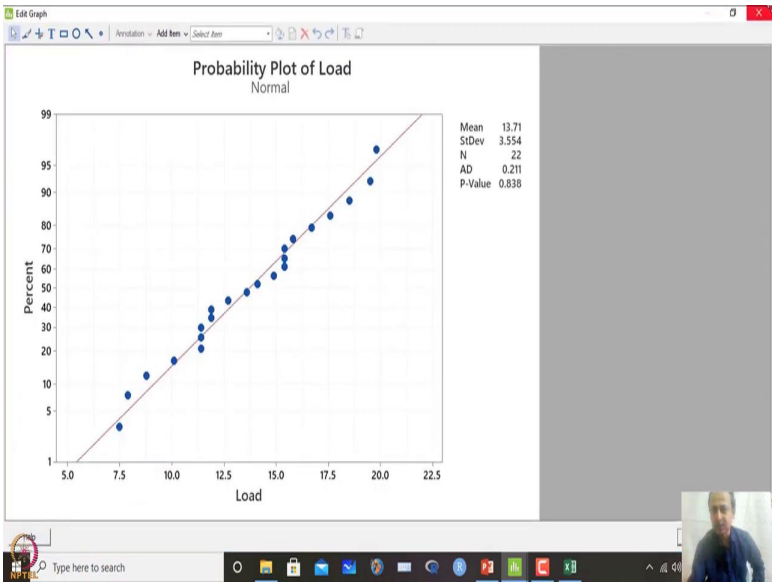
Do the Anderson-Darling test.

(Refer Slide Time: 25:35)



And you will get the information over here.

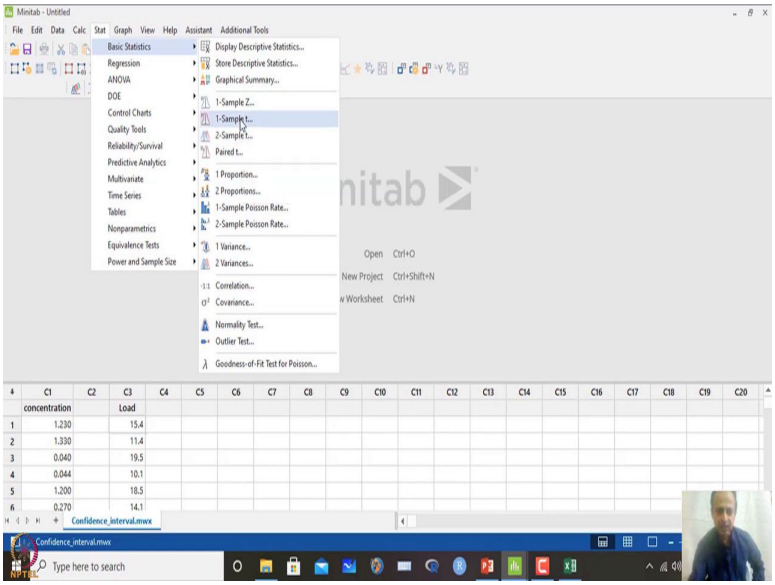
(Refer Slide Time: 25:38)



And, when you see the information over here, you see the p value of the is reflecting that 0.836 . So, it is more than 0.05. So, immediately we can interpret that we can interpret over here that the dataset is adhering to normal distribution assumption. Because P-value is more than 0.05 we have mentioned that if it is more than 0.05 it is normal distribution it is following normal distribution over here.

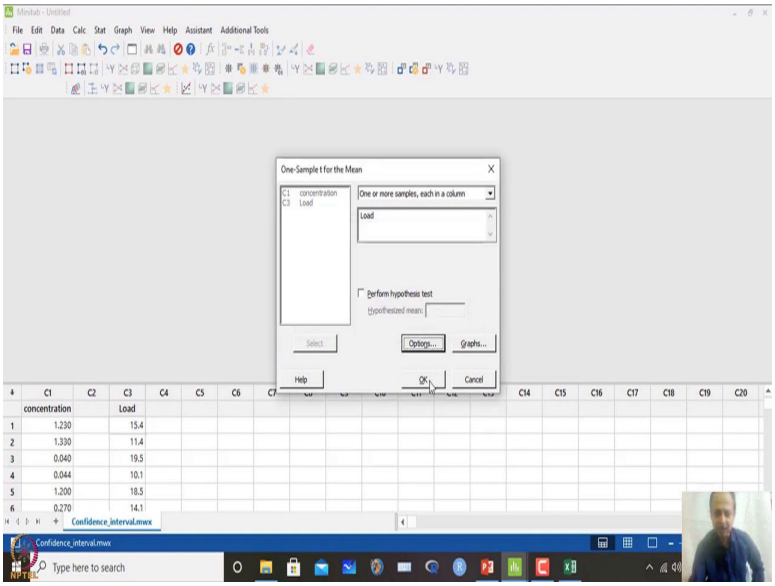
So, this assumption comes out to be true. So, immediately we can use the t statistics over here to calculate confidence interval.

(Refer Slide Time: 26:11)



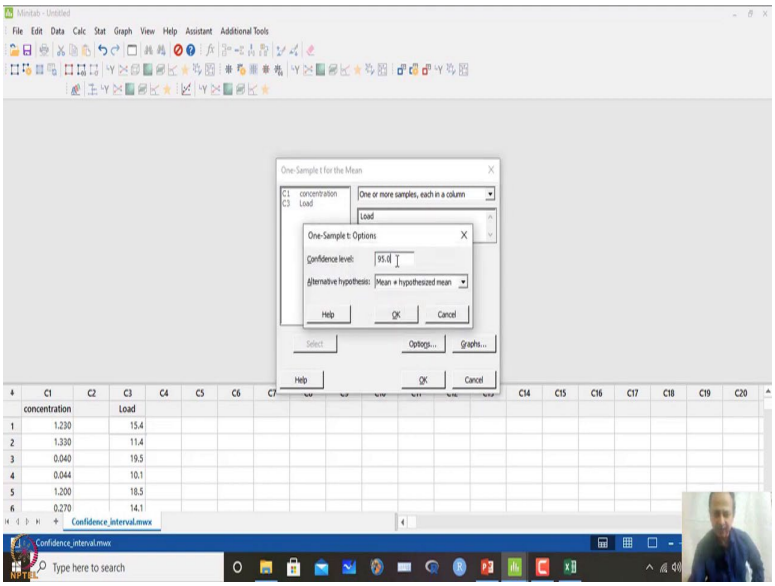
For that what you have to do is that we have to go to Stat, Basic Stat and in that case I have 2 I have 1-sample t over here.

(Refer Slide Time: 26:19)



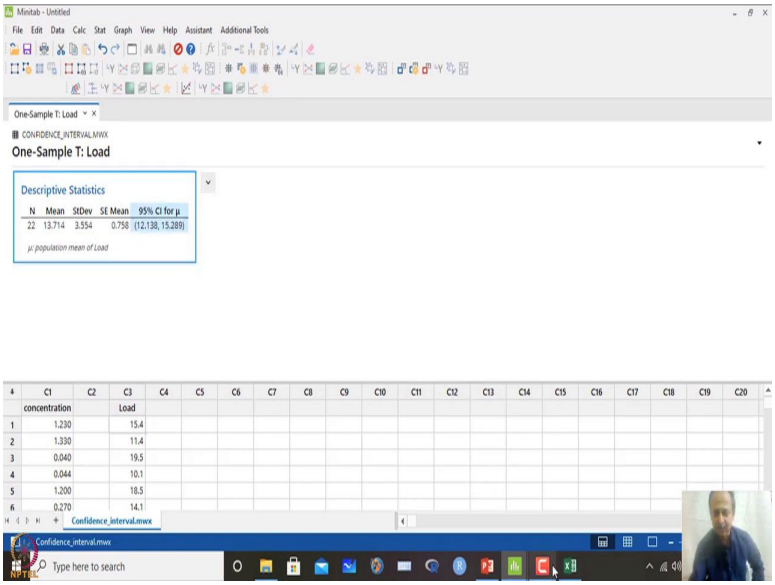
So, instead of this so, I have to mention load as a data information. I will not perform any testing over here.

(Refer Slide Time: 26:25)



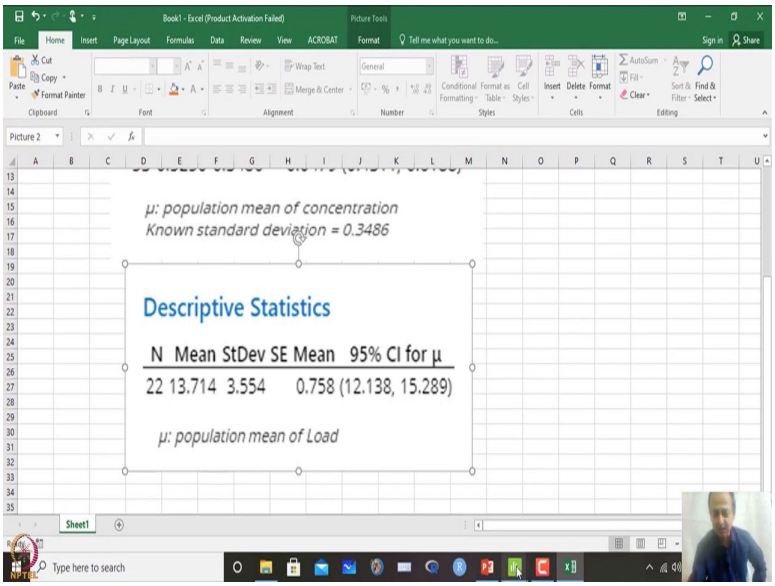
In options, 95 percent confidence interval I am keeping over here. So, I will click Ok over here. And, I will click Ok over here.

(Refer Slide Time: 26:32)



Whenever I do that I have these values of confidence interval. So, Copy as a picture and we can paste it in Excel and try to see what values it is giving like that.

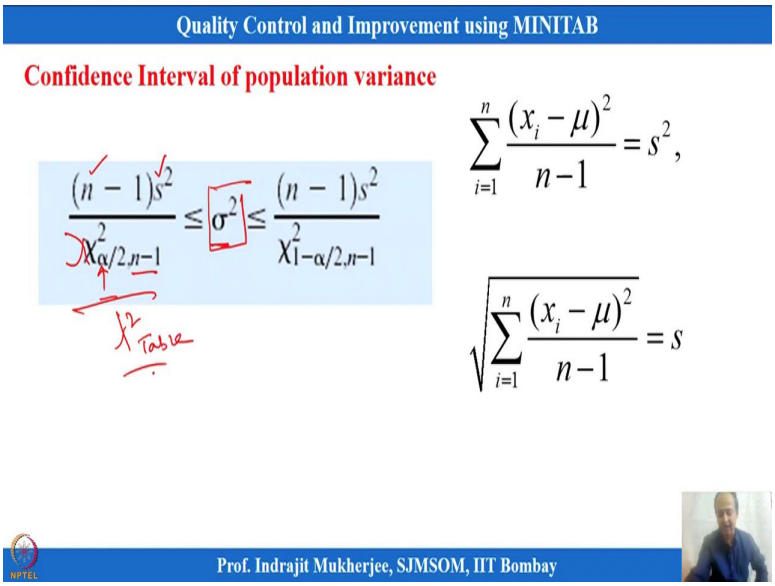
(Refer Slide Time: 26:41)



Earlier it was that is Z distribution, now the interval I am using for these t-distribution over here to calculate. So, 12.138, 15.289 is the values that we are getting over here and let us go back and check what is our calculation. So, 12.14 which is close to 12.138 third place of decimal 15.28, and here also 15.289 like that.

So, MINITAB is also giving the same results when hand calculation what we are doing. So, in this case 95 percent confidence interval MINITAB is giving you directly over here ok. So, another option is if standard deviation is we want to see and similarly, what we can do is that we can also calculate for this, what is confidence interval for the variability of this; that means, variation over here.

(Refer Slide Time: 27:32)



So, what will be the population variance and for that what we can do is that we use as a we use a specific chi square distribution over here. So, in this case what you see is that a chi square distribution can define the confidence interval of σover here for any given data.

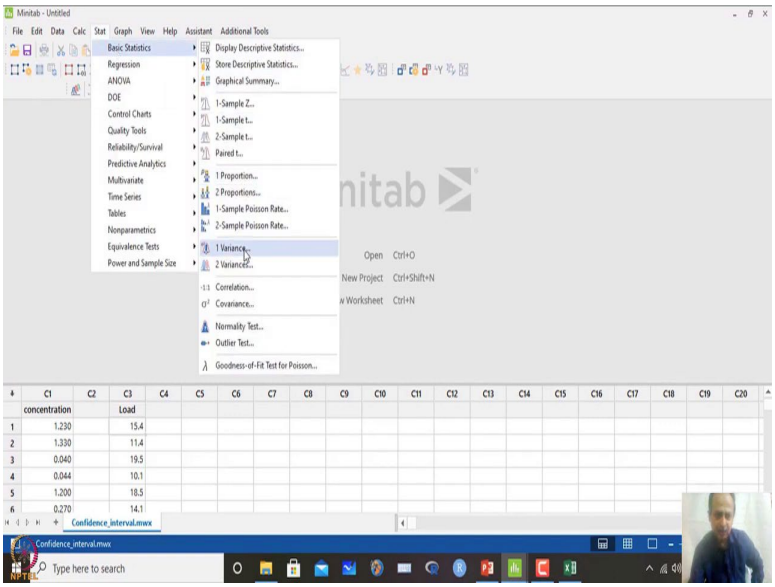
And, the what we need is that s information, number of observations that is selected over here, α values that we are selecting same way and this has to follow a chi square distribution with α/2 and (n-1) degree of freedom like that. So, this is the level of significance over here, α is considered as the level of significance.

So, I can also check the confidence interval of variance over here. So, both the things are possible and here, chi-square distribution which is coming from also normal distribution assumptions basic assumptions and defining a normal another random variables which will follow chi-square.

So, in this case what we are doing is that, they are representing that as a chi-square distribution. So, this statistics this can be can these values can we can get it from chi-square tables like that. So, in this case we have some tables from where we will get the values of chi-square over here.

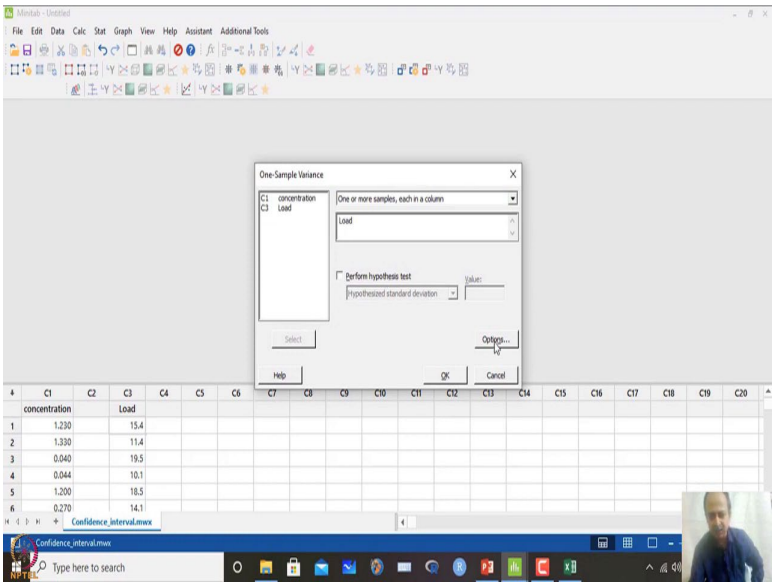
So, I can do that and MINITAB does it automatically for you; for mean also, for standard deviation. So, it will give a confidence interval for the mean also, it will give a confidence interval for the variance also over here.

(Refer Slide Time: 28:56)



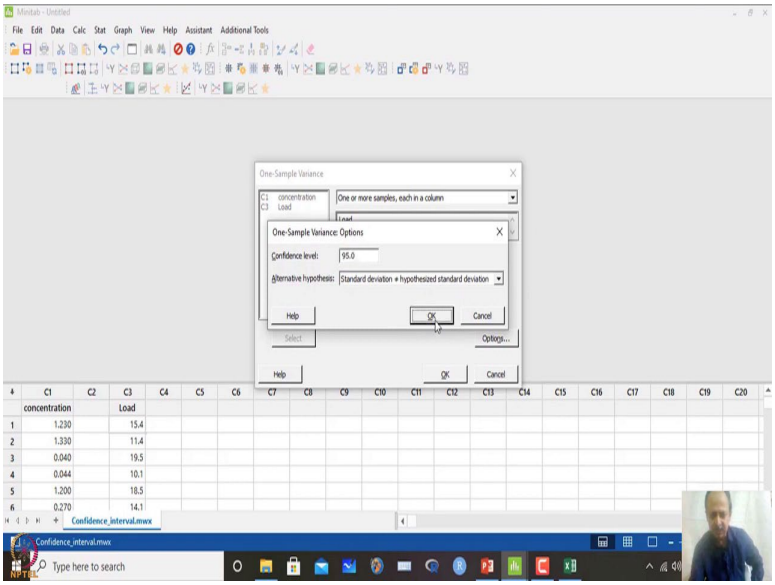
So, if I have to calculate variance confidence interval what I have to do is that I have to go over here and in this case let us go to Stat and Basic Stat and I have 1-Variance like that.

(Refer Slide Time: 28:59)



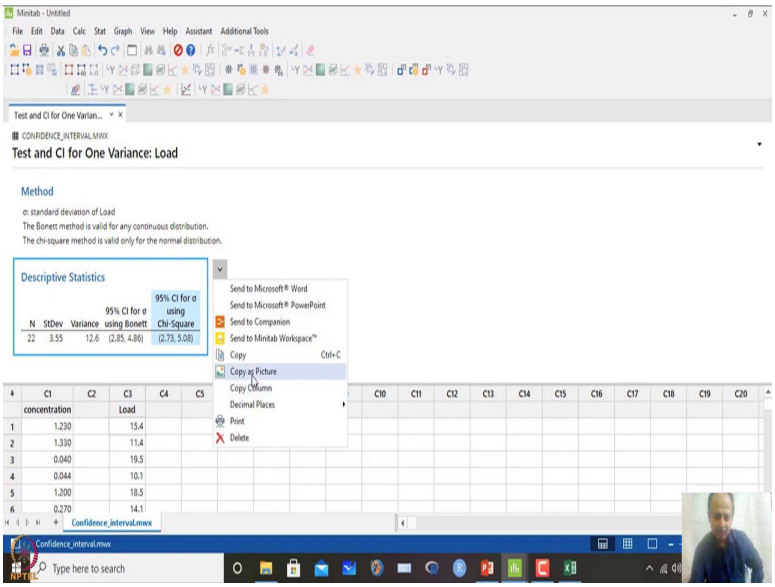
So, in this case I will say Load I want to estimate the variance. I will not perform any hypothesis testing.

(Refer Slide Time: 29:05)



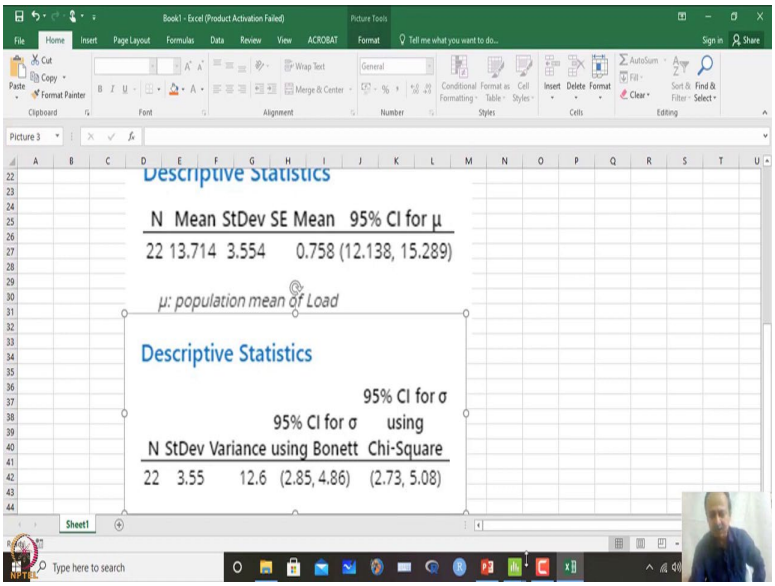
95 percent confidence interval I will keep.

(Refer Slide Time: 29:09)



And, same way if I click Ok over here, so, this is chi square interval that is giving. So, here you can see like that 2 point.

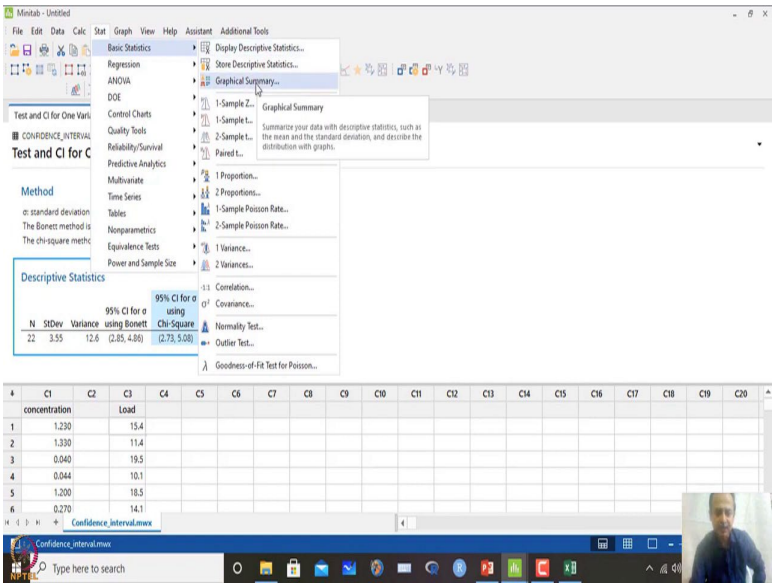
(Refer Slide Time: 29:20)



So, if I copy this as a picture, I can paste this one below this one and what you get is that confidence interval of using the standard process that is chi square distribution though.

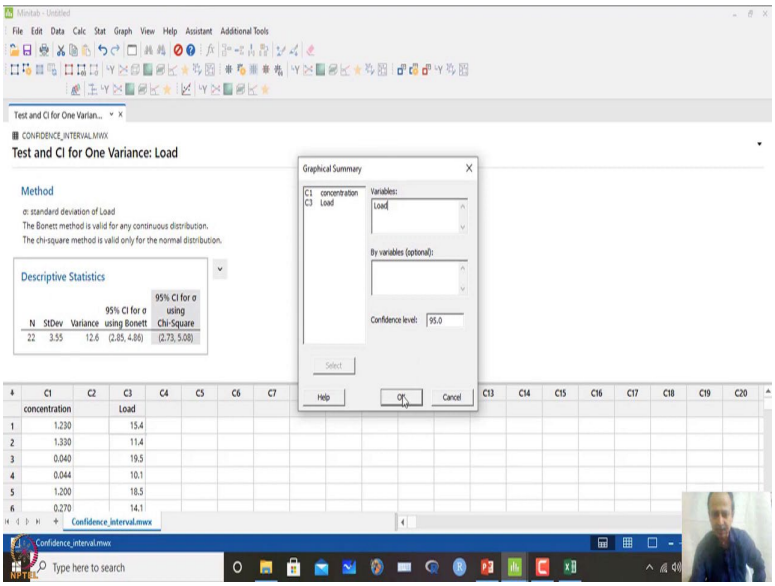
So, you have to see over here 2.73 to 5.08 like that ok, that is possible from here.

(Refer Slide Time: 29:34)



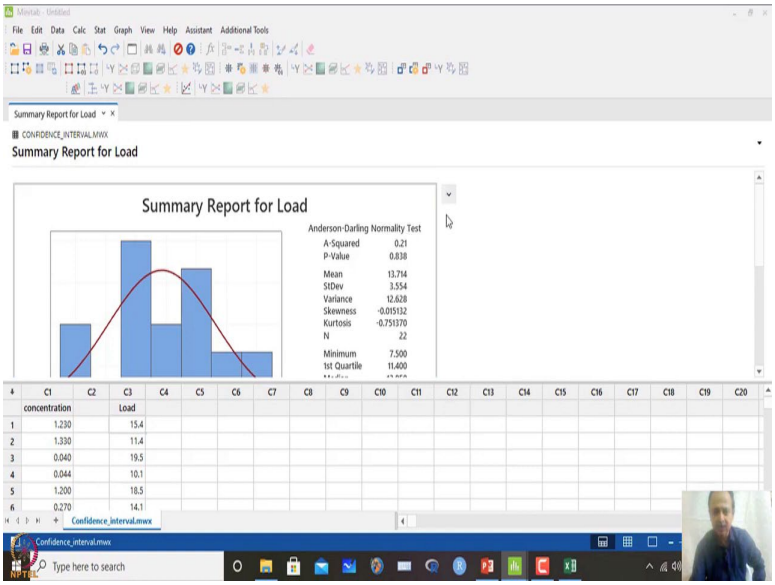
Also there is a graphical summary over here. If you go to basic Stat, graphical summary; if you go to graphical summary what will happen is that you click on graphical summary I want to see for load and confidence level is given as 95 percent.

(Refer Slide Time: 29:40)

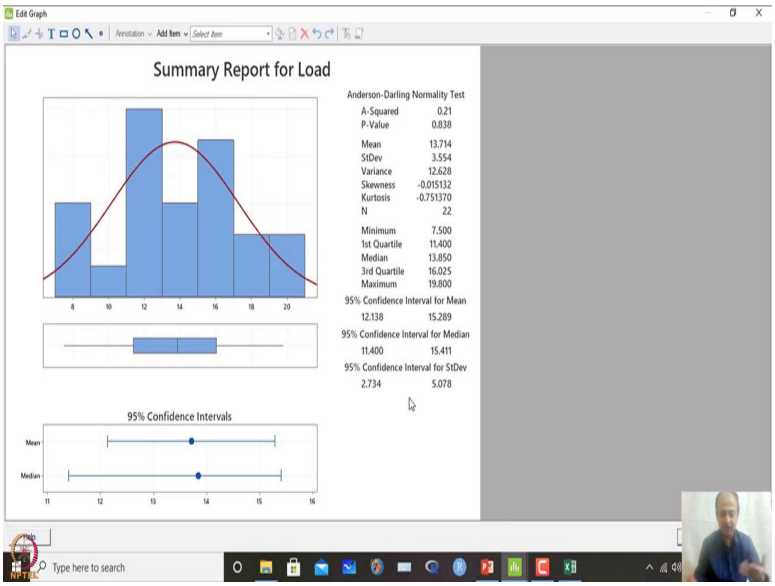


So, if you click Ok, what will happen is that you will get some graphical summary over here.

(Refer Slide Time: 29:48)



(Refer Slide Time: 29:54)



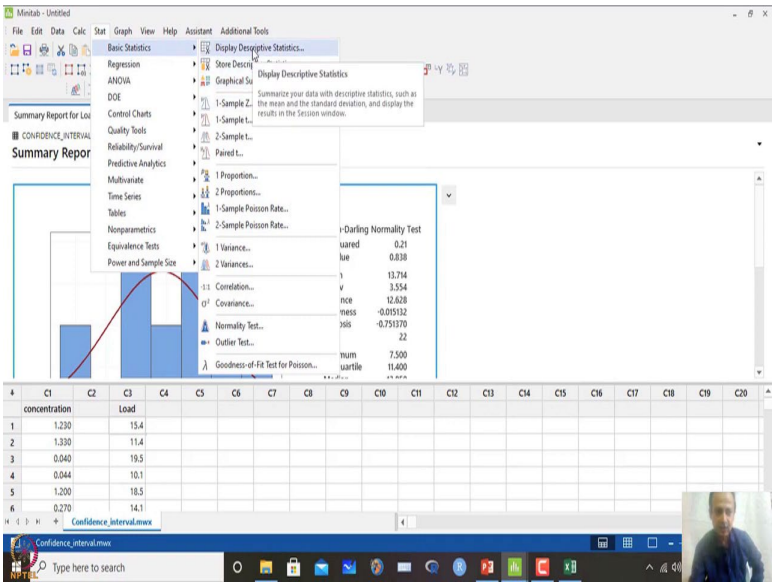
So, in this case what is what you can see is basically, all information together over here all information together over here. So, Anderson-Darling normality test was done and P-value is more than 0.05. So, the data seems to be normal. So, 0.838 is more than P that P value that α value that we have taken as 5 percent like that. So, this is coming out of to be more than 0.05.

So, this is adhering to that mean value, standard deviation, variance, skewness, kurtosis, normal distribution is given, minimum all these statistic, these basic statistic information is given over here. 95 percent confidence interval of mean, so, over here I have not mentioned standard deviation.

So, in this case it will automatically calculate based on t distribution and it can calculate interval for median, it can also calculate interval for standard deviation over here.

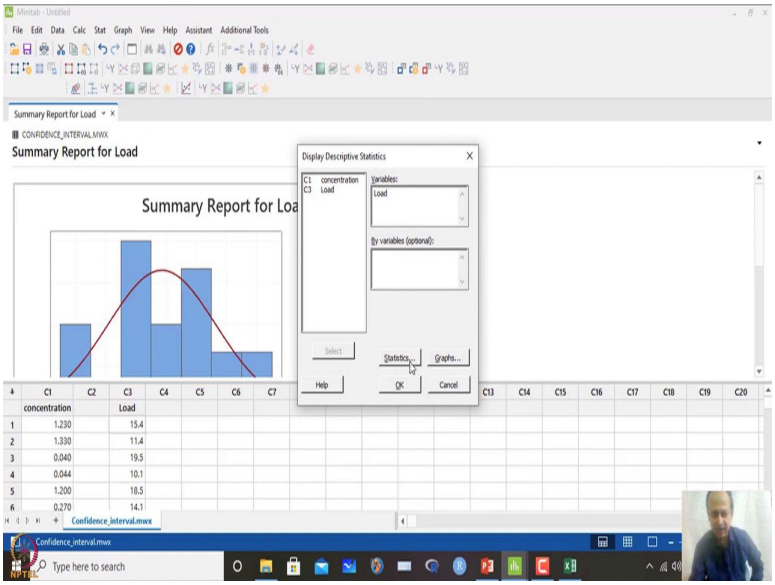
So, in this case only thing is that, what is underlying formula you can see from MINITAB help like that and you can find out, figure out you can do it by hand also; so, confidence interval of mean and standard deviation can also be determined over here, can also be seen over here based on the estimation over here. And, chi-square distribution can be used over here and X-bar in this case t distribution can be used over here ok.

(Refer Slide Time: 31:05)



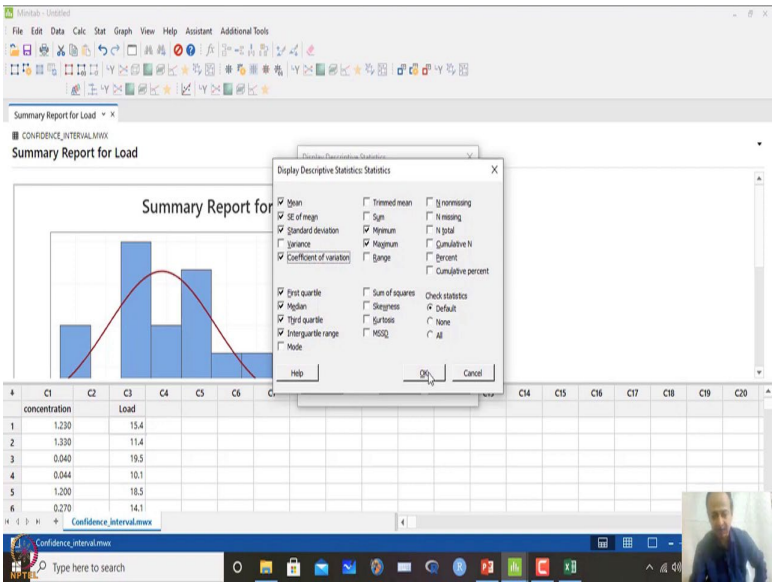
So, I can get it in one go over here, what I have done is that basic statistics graphical summary over here.

(Refer Slide Time: 31:11)



And if you go to descriptive statistics over here, if I give this data over here,

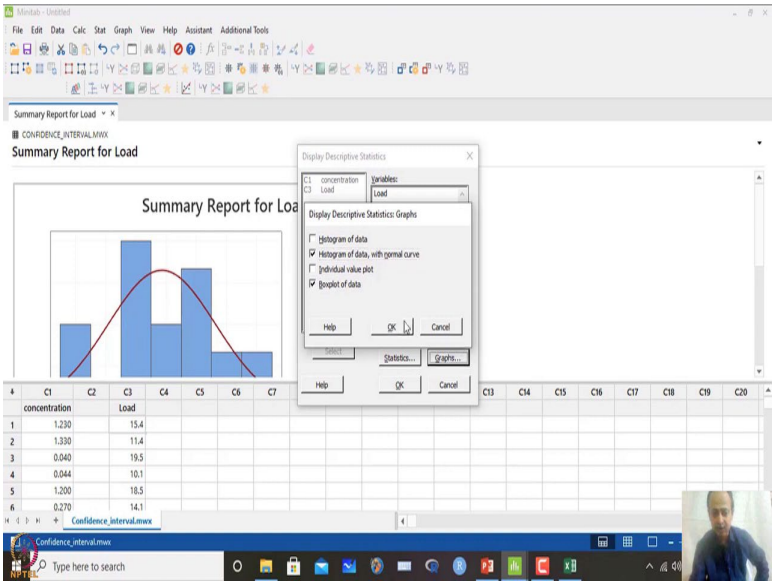
(Refer Slide Time: 31:16)



And, then I go to statistics over here, here also there is possibility that I can see standard error standard deviation, mean, median and loss information no missing variables over here. So, inter quartile range over of the dataset coefficient of variation also we can see which is the ratio between ah σby mean basically the. So, how much is the standard deviation with respect to the mean, how much is the magnitude of standard deviation with respect to the.

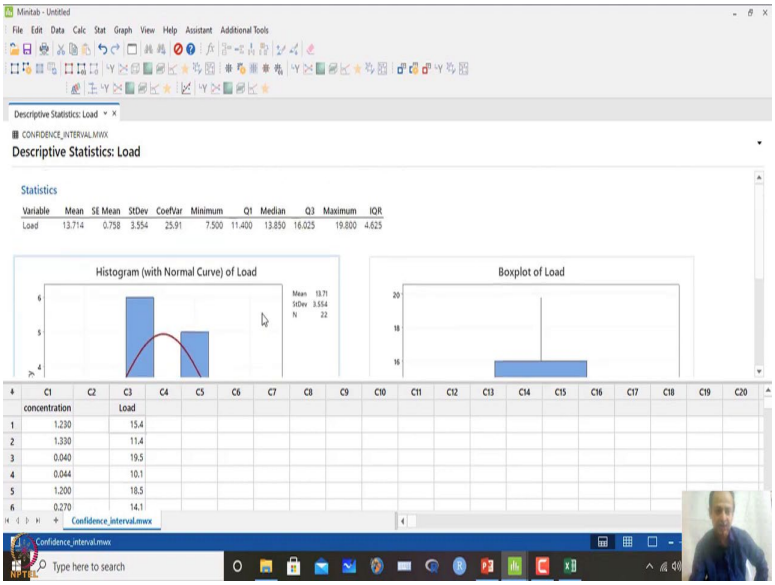
See if I click Ok over here.

(Refer Slide Time: 31:43)



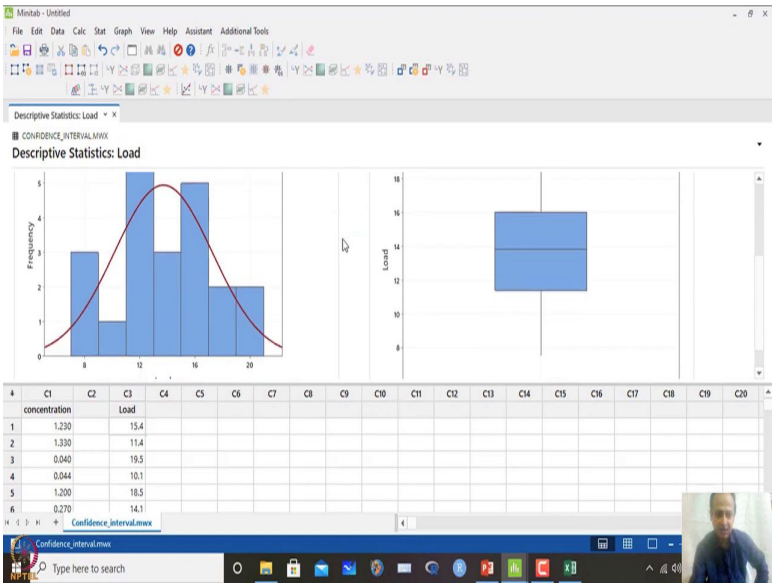
And, in graph also you can see the boxplot of the dataset that is given and Histogram with normal distribution that is also possible over here.

(Refer Slide Time: 31:51)



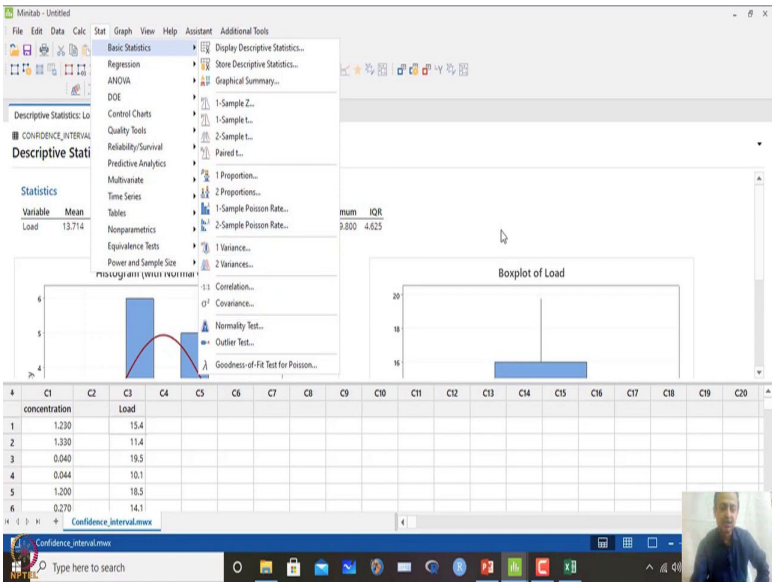
Display descriptive statistics. So, this is the descriptive statistics that you see.

(Refer Slide Time: 31:55)



This is the graph what we have seen earlier and this is the boxplot of the data that it is giving. So, only thing is that over here you will not get the confidence interval.

(Refer Slide Time: 35:05)



So, but that can be seen when we are going to basic stat and we are using graphical summaries over here or otherwise what you can do confidence interval, if σis known I will use 1-Sample Z; if σis unknown in that case I will use 1-Sample t and for variance, I will say 1-Variance test over here.

So, but when I am using t I am assuming the normal distribution assumptions and we can check that point when we implement and try to figure out what is the confidence interval of the mean based on one sample.

So, what we are trying to do or trying to say over here in this session what we have trying to say is that with one single X-bar information and s information we can predict the behavior of the population. We can predict the parameters or interval where μ will lie or σwill lie basically ok.

So, this is a unique thing that is given by the statistician and we are using that, so that we do not have to do experiments time and again like that. In this kind of experiments only with one go, we have to we have to predict basically, one go we have to figure out what should be the setting like that. So, in that case this confidence interval concept is very important.

And, from here we will continue and try to figure out that more statistical information what is required like hypothesis testing that we will try to address in the next lecture ok.

So, thank you for listening. So, in next lecture, we will talk about basics of hypothesis testing which will be used for our experimentation which will be useful. So, I will not discuss huge hypothesis testing concepts like that.

So, I will give you some hints, what is hypothesis testing and based on that we will proceed ok, in our course. So, this is quality control and improvement. So, we this is not a statistics course. So, but we are highlighting what is required in our course so, that way we are going. So, we can stop here and we will continue from here.

Thank you.